Identifying Patients with Major Diabetes-related Complications

(N2C2 Shared Task 2018, Track 1)
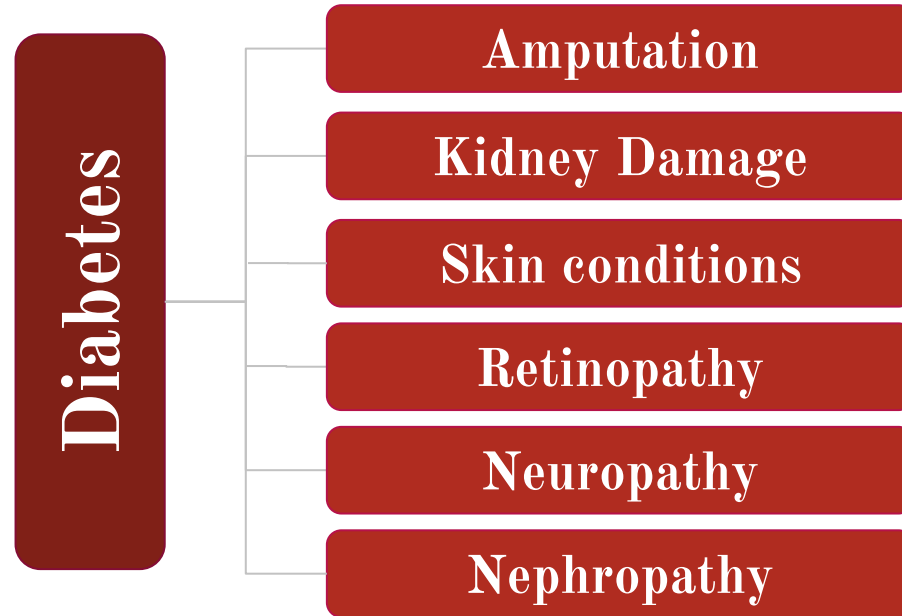
Madhumita Sushil

CLiPS
Computational Linguistics & Psycholinguistics
University of Antwerp

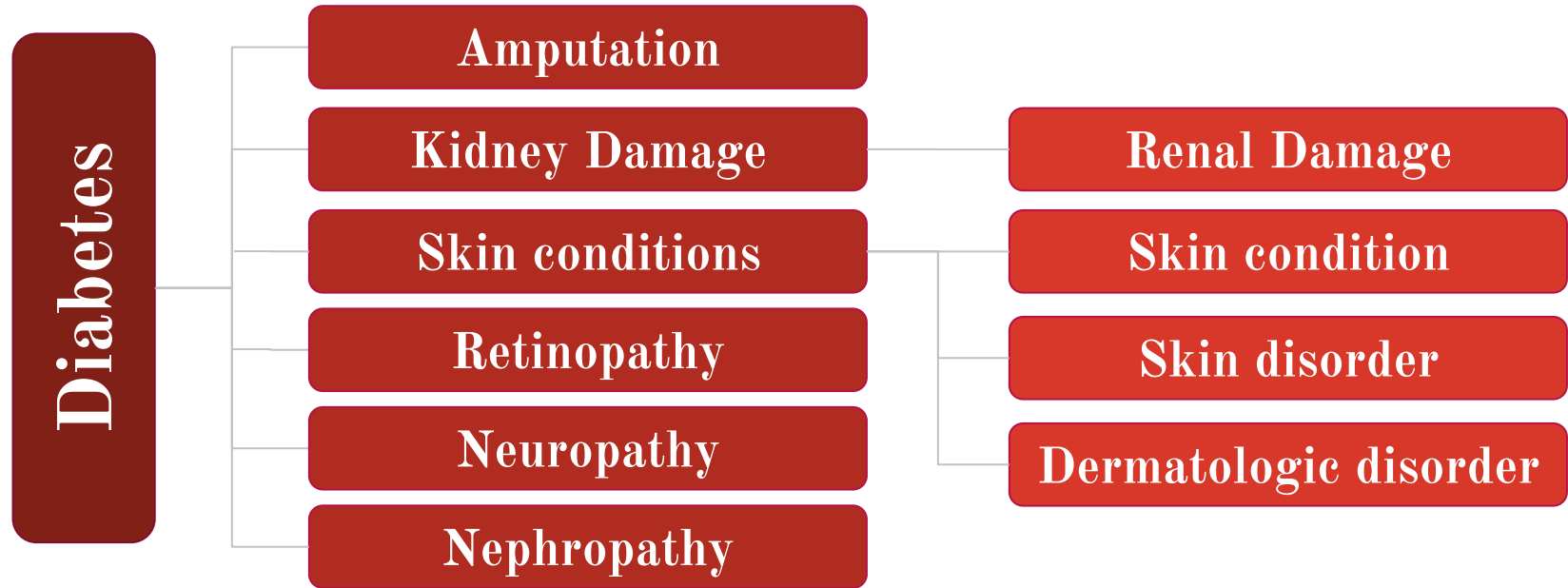# Major diabetes-related complications

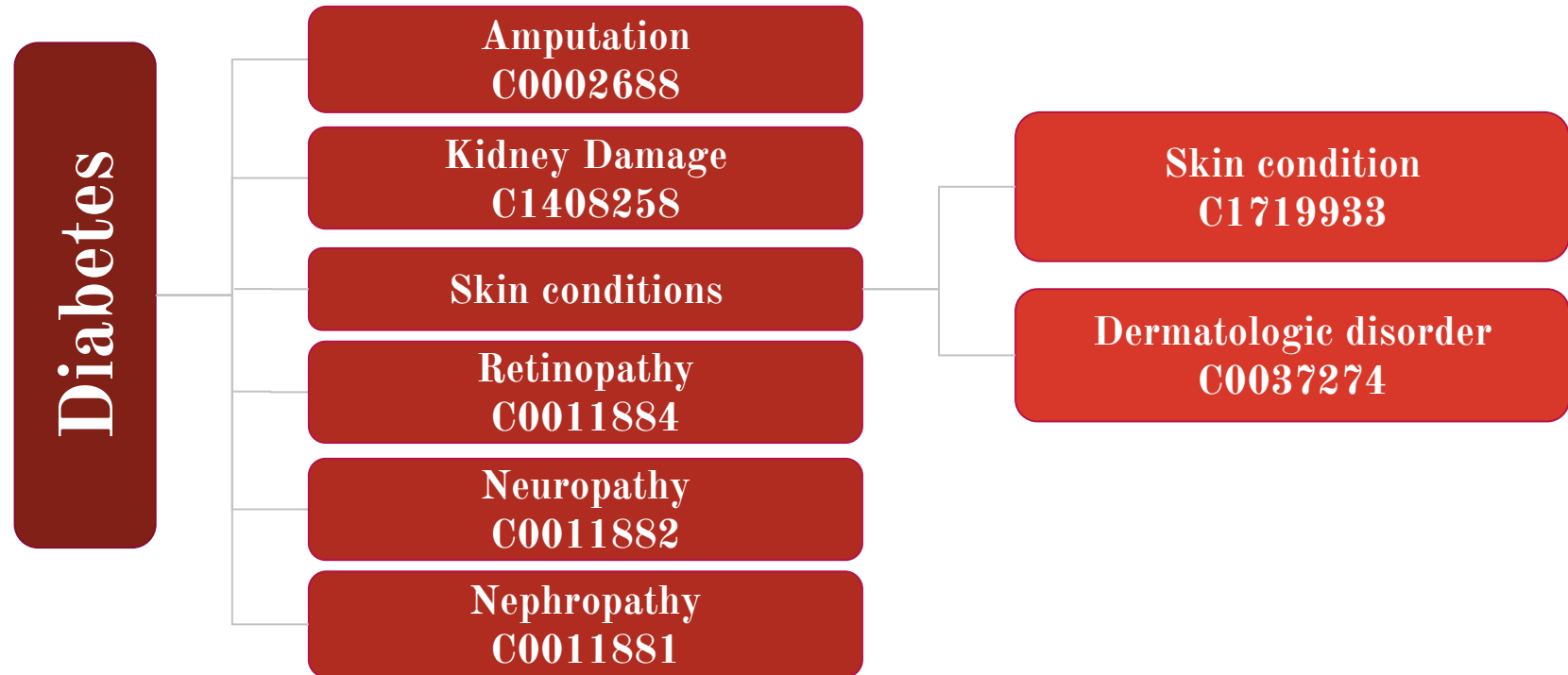# Dataset

Set of ~200 Patients

Has any complication (criteria "**met**")
or
not (criteria "**not met**")

Approaches

Major complications: UMLS extension

Major complications: CUI resolution from UMLS

Diabetes

- **Amputation** C0002688
- **Kidney Damage** C1408258
- **Skin conditions**
  - **Skin condition** C1719933
  - **Dermatologic disorder** C0037274
- **Retinopathy** C0011884
- **Neuropathy** C0011882
- **Nephropathy** C0011881

Embeddings

Word embeddings:

- Word2vec embeddings trained on Pubmed
- **"Complication embeddings"** - average word vectors

CUI embeddings:

- Pretrained CUI embeddings (Beam et al., 2018)

Concept-embeddings

Concept recognition	Concept filtering	Inducing embeddings

1) Mention
2) Semantic type
3) Assertion
4) CUI

Limit to "problems" that are "present"

"Problem" embeddings
1) Average word vector embeddings
2) CUI embeddings

Matching criteria

1) Regex match to "diabetes", "diabetic", and "DM" to check if patient has diabetes, "**not met**" otherwise

2) If diabetic, and for any "present" problem complication pair

$$\cos(\text{prob\_emb}, \text{comp\_emb}) > 0.65 : \text{criteria "\textbf{met}"}$$

Criteria **"not met"** otherwise

Combining CUI and word embedding similarity

Using both approaches and combining them:

1) Max of both similarity scores
2) Average to both similarity scores

Results (train)

| | ----------- met ------------- | | | | ------ not met ------- | | | -- overall --- | |
	Prec.	**Rec.**	**Speci.**	**F(b=1)**	**Prec.**	**Rec.**	**F(b=1)**	**F(b=1)**	**AUC**
Majority	0.5594	1.0000	0.0000	0.7175	0.0000	0.0000	0.0000	0.3587	0.5000
RF - BoW	0.7767	0.7080	0.7416	0.7407	0.6667	0.7416	0.7021	0.7214	0.7248
NB - BoW	0.6827	0.6283	0.6292	0.6544	0.5714	0.6292	0.5989	0.6267	0.6288
RF - BoEmbCluster	0.6194	0.7345	0.4270	0.6721	0.5588	0.4270	0.4841	0.5781	0.5807
Word emb sim	0.8404	0.6991	0.8315	0.7633	0.6852	0.8315	0.7513	0.7573	0.7653

More Results (train)

	----------- met ------------				------ not met -------			-- overall ---	
	Prec.	Rec.	Speci.	F(b=1)	Prec.	Rec.	F(b=1)	F(b=1)	AUC
Word emb sim	0.8404	0.6991	0.8315	0.7633	0.6852	0.8315	0.7513	0.7573	0.7653
CUI similarity	0.8356	0.5398	0.8652	0.6559	0.5969	0.8652	0.7064	0.6812	0.7025
Max similarity	0.7807	0.7876	0.7191	0.7841	0.7273	0.7191	0.7232	0.7537	0.7534
Avg similarity	0.8020	0.7168	0.7753	0.7570	0.6832	0.7753	0.7263	0.7417	0.7460